

Better Online Deterministic Packet Routing on Grids

Guy Even *

Moti Medina [†]

Boaz Patt-Shamir ^{‡ §}

January 27, 2015

Abstract

We consider the following fundamental routing problem. An adversary inputs packets arbitrarily at sources, each packet with an arbitrary destination. Traffic is constrained by link capacities and buffer sizes, and packets may be dropped at any time. The goal of the routing algorithm is to maximize throughput, i.e., route as many packets as possible to their destination. Our main result is an $O(\log n)$ -competitive deterministic algorithm for an n -node line network (i.e., 1-dimensional grid), requiring only that buffers can store at least 5 packets, and that links can deliver at least 5 packets per step. We note that $O(\log n)$ is the best ratio known, even for randomized algorithms, even when allowed large buffers and wide links. The best previous deterministic algorithm for this problem with constant-size buffers and constant-capacity links was $O(\log^5 n)$ -competitive. Our algorithm works like admission-control algorithms in the sense that if a packet is not dropped immediately upon arrival, then it is “accepted” and guaranteed to be delivered. We also show how to extend our algorithm to a polylog-competitive algorithm for any constant-dimension grid.

Keywords. Online Algorithms, Packet Routing, Bounded Buffers, Admission Control, Grid Networks

1 Introduction

The core function of any packet-switching network is to route packets from their origins to their destinations, but many fundamental questions about packet routing are far from being well understood. In this paper we consider one of these questions, namely the *competitive throughput network model*, introduced by [3].

Briefly, the model is as follows. The network consists of n nodes (switches) connected by point-to-point unidirectional communication links, and we are given two positive integer parameters, B and c , called the buffer size and link capacity, respectively. Executions proceed as follows. Packets are input by an adversary over time. Each packet is input at its *source* node with a given *destination* node. At each step, each packet is either forwarded over an incident link, stored in its current location buffer, or dropped (i.e., removed from the system). Storing and forwarding are subject to the constraints that a buffer can store at most B packets simultaneously and that a link can carry at most c packets in a time step. These constraints can be met for all input sequences since the model allows for packets to be dropped at any time. The *routing algorithm* selects, at each step, which packets are forwarded, which are stored, and which are dropped. The goal of the algorithm is to maximize the number of packets delivered at their destination. Since we consider

*School of Electrical Engineering, Tel Aviv Univ., Tel-Aviv 69978, Israel. (guy@eng.tau.ac.il).

[†]LIAFA, Université Paris Diderot, 75205 Paris Cedex 13, France. (moti.medina@liafa.univ-paris-diderot.fr).

[‡]School of Electrical Engineering, Tel Aviv Univ., Tel-Aviv 69978, Israel. (boaz@tau.ac.il).

[§]Supported in part by ISF, MoST and Neptune.

Ref.	Dim.	Comp. Ratio	Deterministic?	Range of B, c	Remarks
[11, 12, 13]	1	$O(\log n)$	✓	$B, c > \log n, B/c = n^{O(1)}$	immediate from s-t reduction
[4]	1	$O(\log^3 n)$	—	$B \geq 2, c = 1$	
[7]	1	$O(\log^2 n)$	—	$B \geq 2, c = 1$	FIFO buffers
[11, 13]	1	$O(\log n)$	—	$B \in [1, \log n], c \geq 1$	also for $\log n \leq B/c \leq n^{O(1)}$
[12, 13]	1	$O(\log^5 n)$	✓	$[3, O(\log n)]$	preemptive
[12, 13]	d	$O(\log^{d+4} n)$	✓	$[3, O(\log n)]$	preemptive
Theorem 11	1	$O(\log n)$	✓	$[5, O(\log n)]$	
Theorem 19	d	$2^{O(d)} \cdot \log^d n$	✓	$[2^{d+1} + 1, O(\log n)]$	

Table 1: Some results for centralized online algorithms for packet routing. The networks are uni-directional grids. In the special case of $B = 0$ and $c \geq 3$, the algorithm in [12, 13] is $O(\log^{d+2} n)$ -competitive.

on-line algorithms, we evaluate algorithms by their competitive ratio, i.e., the minimum ratio, over all finite packet input sequences, between the number of packets delivered by the on-line algorithm and the maximum number of packets that can be delivered by any (off-line) constraint-respecting schedule.

It is nearly an embarrassment to find that very little is known about this problem, even in the simplest case, where the network topology is the trivial n -node unidirectional line. In this work we provide an improved deterministic algorithm for networks whose topology is a d -dimensional grid.

1.1 Our Results

Our main result is a centralized *deterministic* $O(\log n)$ -competitive packet routing algorithm for unidirectional lines with n nodes. The algorithm requires buffer size $B \geq 5$ and link capacity $c \geq 5$. In addition, both B and c must be $O(\log n)$. We show how to extend the algorithm to d dimensions, where the competitive ratio is $2^{O(d)} \cdot \log^d n$, assuming that $B, c \geq 2^{d+1} + 1$. Our algorithm is *nonpreemptive*, namely, packets are dropped only at the time of their arrival (similarly to admission control policies, which “accept” or “reject” requests upon arrival). By contrast, preemptive algorithms may drop packets at any time, i.e., packets are not guaranteed to reach the destination even after they start traversing the network. The best previous deterministic algorithm [12, 13] is preemptive.

Table 1 provides a summary of our results and a comparison of our algorithm with some previous results along various aspects.

1.2 Overview of Techniques

We first explain our approach for the 1-dimensional case.

The high-level idea is to reduce packet routing in a graph G to circuit switching (or path packing, see [15, 6]) in the *space-time graph* $G \times T$, where T denotes the set of time steps. This so-called space-time transformation has been used extensively in this context [5, 2, 7, 16, 11, 12, 13]. To be effective, the space-time transformation requires an upper bound on path lengths which does not result in losing too much throughput. We use the bound of [12, 13] (which extends [7]), that ensures that the loss is at most some constant fraction. After the transformation, we have an instance of *online path packing* [6, 9]. It is known that if the capacities are large enough, i.e., $\log n \leq B, c \leq n^{O(1)}$, then online path packing is solvable with logarithmic competitive ratio [6, 11, 12, 13]. We overcome the difficulty that B and c are $O(\log n)$ by employing a technique called *tiling*, i.e., partitioning the network nodes into large enough subgrids. Tiling has been used in the past [15, 8, 11, 12, 13]; in our algorithm, we use 4 distinct tilings, and work on each of them independently. Each tiling induces a new graph called the *sketch graph* whose nodes are the tiles. The capacity of the edges in the space-time graph between adjacent tiles is $O(\log n)$ to allow for applying $O(\log n)$ -competitive path packing algorithms. Path packing algorithms over the sketch graph produce sketch paths for accepted packets. Thus, after these preliminary simplifications, we arrive at the sub-task of *detailed routing*, in which coarse sketch paths must be expanded to paths in the original space-time graph.

Fractional Optimum. Key to our application of the path-packing algorithm is the analysis of Buchbinder and Naor [9, 10], which bounds the performance of the algorithm w.r.t. the *fractional* optimum, which may deliver packet fractions. This result allows us to scale buffer sizes and link capacities up and down while keeping the competitive ratio under control.

Combining algorithms. Another central component in the analysis of our algorithm is the combination technique introduced by Kleinberg and Tardos [15]. Loosely speaking, this technique deals with an admission control algorithm that is the conjunction of two competitive algorithms, the state of which depends only on the requests accepted by both. The technique enables one to prove that the competitive ratio of the combined algorithm is the sum (rather than the product) of the competitive ratios of the constituent algorithms.

1.3 Previous Work

Algorithms for dynamic routing on networks with bounded buffers have been studied extensively both in theory and in practice (see, e.g., [1] and references therein). Let us first focus on centralized algorithms for d -dimensional grids. We note that while centralized algorithms for packet routing were always relevant for switch scheduling, recently the idea of centralization of network functions, including route computation, gained substantial additional traction due to the concept of software-defined networks (SDN). See, e.g., [14]. The special case of 2-dimensional grids (with or without buffers) is of particular interest as this is the underlying topology of crossbars in switches [17].

Online Algorithms for Unidirectional Lines. There is a series of papers on uni-directional line networks, starting with [3], which introduced the model. In [3], a lower bound of $\Omega(\sqrt{n})$ was proved for the greedy algorithm on unidirectional lines if the buffer size $B \geq 2$. For the case $B = 1$ (in a slightly different model), an $\Omega(n)$ lower bound for any deterministic algorithm was proved by [7, 4]. Both [7] and [4] developed, among other things, online randomized centralized algorithms for uni-directional lines with $B \geq 2$. In [4] an $O(\log^3 n)$ -competitive randomized centralized algorithm was presented for $B \geq 2$. In addition, it is proved in [4] that nearest-to-go is $\tilde{O}(\sqrt{n})$ -competitive for $B \geq 2$. For the case $B = 1$, [4] presented a randomized $\tilde{O}(\sqrt{n})$ -competitive distributed algorithm. (This algorithm also applies to rooted trees when

all packet are destined at the root.) In [7], an $O(\log^2 n)$ -competitive randomized algorithm was presented for the case $B \geq 2$. (This algorithm also applies to rings and trees.) In [11], an $O(\log n)$ -competitive, nonpreemptive, randomized algorithm was presented. The algorithm in [11] is applicable to a wide range of buffer sizes and link capacities, including the case $B = c = 1$. In [12], an $O(\log^5 n)$ -competitive deterministic algorithm was presented. The algorithm in [12] is applicable for $B, c \in [3, \log n]$.

Online Algorithms for Unidirectional Grids. Angelov et al. [4] showed that the competitive ratio of greedy algorithms in unidirectional 2-dimensional grids is $\Omega(\sqrt{n})$ and that nearest-to-go policy achieves a competitive ratio of $\tilde{\Theta}(n^{2/3})$. In [12], an $O(\log^6 n)$ -competitive deterministic algorithm was presented. An extension of this algorithm to d -dimensional unidirectional grids, with competitive ratio $O(\log^{d+4} n)$, is presented in [12].

For more related results, refer to [13].

Organization. The problem is formalized in Section 2. In Section 3 we explain the reduction of packet-routing to path packing, and the construction of sketch graph. In Section 4 we describe the overall algorithm, and in Section 5 we analyze it. Sections 3–5 deal with the 1-dimensional grid (line); extension to the d dimensional case is also discussed in Section 6.

2 Model and Problem Statement

We consider the standard model of synchronous store-and-forward packet routing networks [3, 4, 7]. The network is modeled by a directed graph $G = (V, E)$, and by two integer parameter $B, c > 0$. For the most part of this paper, we consider a network whose topology is a *directed line* of n vertices, i.e., $V = \{v_0, \dots, v_{n-1}\}$, $E = \{(v_{i-1}, v_i) \mid 0 < i < n\}$.

Execution proceeds in discrete steps. In step t , an arbitrary set of *requests* is input to the algorithm. Each request represents a packet, and we will use both terms interchangeably. A request is specified by a 3-tuple $r_i = (a_i, b_i, t_i)$, where $a_i \in V$ is the *source node* of the packet, $b_i \in V$ is its *destination node*, and $t_i \in \mathbb{N}$ is the time step in which the request is input.

In each time step, the *routing algorithm* removes packets that reached their destination, and decides, for each packet currently in the network, including packets input in the current step, whether (i) to drop the packet, or (ii) to send it over an incident link, or (iii) to store it in the current node. The selection of the action is done subject to the following considerations.

- If a packet is dropped, it is lost forever.
- A packet sent from node u over link (u, v) at time t will be located at node v at time $t + 1$. The *link capacity constraint* asserts that at any step, at most c packets can be sent over each link.
- A packet stored at node u at time t will be located at node u at time $t + 1$. The *buffer capacity constraint* asserts that at any step, at most B packets can be stored in each buffer.

We use the following terminology. A packet $r_i = (a_i, b_i, t_i)$ is said to be *input* (or *arrive*) at a_i at time t_i . We say that r_i is *rejected* if it is dropped at time t_i , otherwise it is *accepted*. (Our algorithm will guarantee that all accepted packets arrive at their destination.)

Given a set of requests, the *throughput* of a packet routing algorithm is the number of packets that are delivered to their destination. We consider the problem of maximizing the throughput of an online centralized deterministic packet-routing algorithm. By *online* we mean that by time t , the algorithm received as input only requests that have been input by time t . By *centralized* we mean that the algorithm receives all requests and controls all packets currently in the system without delay. By *nonpreemptive* we mean that every accepted packet reaches its destination.

Competitive Ratio. Let σ denote an input sequence. Let ALG denote a packet-routing algorithm. Let $\text{ALG}(\sigma)$ denote the throughput obtained by ALG on input σ . Let $\text{OPT}(\sigma)$ denote the largest possible subset of requests in σ that can be delivered without violating the capacity constraints. We say that an online deterministic ALG is ρ -competitive, if for every input sequence σ , $|\text{ALG}(\sigma)| \geq \frac{1}{\rho} \cdot |\text{OPT}(\sigma)|$. Our goal is to design an algorithm with the smallest possible competitive ratio.

3 First Steps

In this section we present preliminary simplifications we apply to the problem. They include reducing the packet routing on a line problem to path packing on grids, and then path packing on sketch graphs.

3.1 From Packet-Routing on a Line to Path Packing in a Grid

Let $G = (V, E)$ denote a directed line with link capacities c and buffer sizes B . The space-time grid of $G = (V, E)$ is a directed acyclic infinite graph $G^{st} = (V^{st}, E^{st})$ with edge capacities $c^{st}(e)$, where (i) $V^{st} \triangleq V \times \mathbb{N}$. Each vertex $v \in V$ has infinitely many copies in the space-time grid G^{st} ; namely, vertex $(v, t) \in V^{st}$ is the copy of v that corresponds to time t . (ii) $E^{st} \triangleq E_0 \cup E_1$ where E_0 denotes forward edges and E_1 denotes the store edges. Formally, $E_0 \triangleq \{(u, t) \rightarrow (v, t+1) : (u, v) \in E, t \in \mathbb{N}\}$ and $E_1 \triangleq \{(u, t) \rightarrow (u, t+1) : u \in V, t \in \mathbb{N}\}$. (iii) The capacity of all edges in E_0 is c , and all edges in E_1 have capacity B .

The transformation. We transform a request $r_i = (a_i, b_i, t_i)$ for routing a packet in the directed line G to a path request $r_i^{st} = ((a_i, t_i), \text{row}(b_i))$ in the grid G^{st} . The correctness of the reduction is based on a one-to-one correspondence between paths in G^{st} and a routing of a packet in G . Each vertical edge $(v_i, t) \rightarrow (v_{i+1}, t+1)$ in G^{st} corresponds to forwarding a packet from v_i to v_{i+1} in step t , and each horizontal edge $(v_i, t) \rightarrow (v_i, t+1)$ in G^{st} corresponds to storing a packet in v_i in step t .

Embedding in the plane. The naïve depiction of G^{st} maps vertex (v_i, t) to the point (t, i) in the plane (i.e., the x -axis is the time axis and the y -axis is the “vertex-index” axis). This embedding of G^{st} results with a lattice of vertices in which edges are either horizontal or diagonal. We prefer the embedding in which the edges are axis parallel, which means that vertex (v_i, t) is mapped to the point $(t - i, i)$. In the axis-parallel depiction, all the copies of a vertex $v_i \in V$ still reside in the i th row. However, column j corresponds to a traversal of the complete line, starting at v_0 at time j and ending at v_{n-1} at time $j + n - 1$.

3.2 From One Grid to Four Sketch Graphs

Given a grid generated by the transformation above, we apply another transformation to produce a coarsened version, called the *sketch graph*. Specifically, we use *tiling*. Tiling is a partition of the grid nodes into $\ell_h \times \ell_v$ subgrids, where ℓ_h and ℓ_v are parameters to be determined later. We also add dummy nodes to the space-time grid G^{st} to complete all tiles. This augmentation has no effect on routing because a dummy vertex does not belong to any route between real vertices.

The tiling is specified by two additional parameters ϕ_x and ϕ_y called *offsets*. The offsets determine the positions of the corners of the tiles; namely, the left bottom corner of the tiles are located in the points $(\phi_x + i \cdot \ell_h, \phi_y + j \cdot \ell_v)$, for $i, j \in \mathbb{N}$. The algorithm uses four offsets $(\phi_x, \phi_y) \in \{-\ell_h/2, 0\} \times \{-\ell_v/2, 0\}$. We denote these four tilings by T_1, \dots, T_4 .

Proposition 1. *For every vertex (v, t) of the space-time grid G^{st} , there exists exactly one tiling T_j such that (v, t) is in the south-west quadrant of a tile of T_j .*

Proposition 1 suggests a partitioning of the requests.

Definition 2. A request $r_i = (a_i, b_i, t_i)$ is in SW_j if the source vertex (a_i, t_i) of request r_i belongs to the south-west quadrant of a tile in the tiling T_j .

The Sketch Graphs. Each tiling T_j induces a grid, called the sketch graph, each vertex of which corresponds to a tile. The sketch graph induced by T_j is denoted by $S_j \triangleq (V(S_j), E(S_j))$, where $V(S_j)$ is the set of tiles in T_j . There is a directed edge $(s_1, s_2) \in E(S_j)$ if $s_1 \neq s_2$ and $E^{st} \cap (s_1 \times s_2) \neq \emptyset$. All edges in the sketch graph are assigned unit capacity.

3.3 Online Packing of Paths

We use the sketch graphs to solve *path packing* problems. Intuitively, the path packing model resembles the packet routing model, except that there are no buffers, and that each link e may have a different capacity $c(e)$. In addition, we generalize the notion of a request to allow for a set of destinations (similar to “anycast”) as follows. Usually, the destination of a request consists of a single vertex. If G is a directed graph, then it is easy to reduce the case in which the destination is a subset to the case in which the destination is a specific vertex. The reduction simply adds a sink node that is connected to every vertex in the destination subset. In our setting of space-time grid, the destination subset is a row. Thus it suffices to add a sink node for each row (as in [7]).

Formally, a *path request* r_i in G is a pair (a_i, D_i) , where $a_i \in V$ is the source vertex and $D_i \subseteq V$ is the destination subset. Let $P(r_i)$ denote the set of paths that can be used to serve request r_i ; namely, every path $p \in P(r_i)$ begins in a_i , ends in a vertex in D_i , and satisfies some additional constraint (e.g., bounded length, bounded number of turns, etc.). Given a sequence $R = \{r_i\}_{i \in I}$ of path requests, we call a sequence $P = \{p_i\}_{i \in J}$ a *partial routing* of R if $J \subseteq I$ and $p_i \in P(r_i)$ for every $i \in J$. The *load* of an edge $e \in E$ induced by P is the ratio $\frac{|\{p_j \in P: e \in p_j\}|}{c(e)}$. A partial routing of a set of path requests is called a β -*packing* if the load induced on each edge is at most β . The *throughput* of P is simply the number $|J|$ of paths in P .

Integral and Fractional Partial Routings. In the integral scenario, a path request is either served by a single path or is not served. In *fractional routing*, a request r_i can be (partially) served by a combination of paths p_1, \dots, p_k . Namely, each path p_j serves a fraction λ_j of the request, where $\lambda_j \geq 0$ for all j and $\sum_j \lambda_j \leq 1$. We refer to $\sum_j \lambda_j$ as the *flow amount* of request r_i . The *load* of an edge $e \in E$ induced by request r_i is the ratio $\sum_{j: e \in p_j} \lambda_j / c(e)$. A fractional solution is β -packing if the total load on in each edge, from all requests, is at most β . The *throughput* of a fractional routing is the sum of the flow amounts of all requests. Given a fractional routing g , we use $|g|$ to denote its throughput. Trivially, the maximum throughput attainable by a fractional β -packing is an upper bound on the maximum throughput attainable by an integral β -packing. An optimal-throughput fractional β -packing can be computed off-line by solving a linear program.

Online Path Packing: Problem and Solution. In the online path packing problem, the input is a sequence of path requests $R = \{r_i\}_{i \in I}$. Upon arrival of a request r_i , the algorithm must either allocate a path $p \in P(r_i)$ to r_i or reject r_i . An online path packing algorithm is said to be (α, β) -*competitive* if it computes a β -packing whose throughput is at least $1/\alpha$ times the maximum throughput over all 1-packings. Note that for online path packing, we assume that all edges have capacity at least 1.

The online path packing algorithm in [6] (analyzed also by [9]) assigns weights to the edges that are exponential in the load of the edges. This load is the load incurred by the paths allocated to the requests

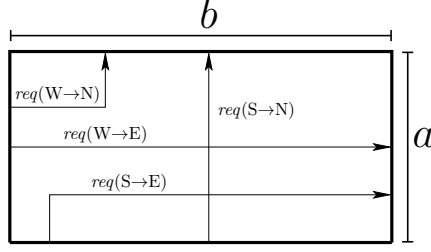


Figure 1: An $a \times b$ grid. The four types of requests (i.e., $\text{req}(W \rightarrow N)$, $\text{req}(S \rightarrow N)$, $\text{req}(W \rightarrow E)$, $\text{req}(S \rightarrow E)$) are depicted by the four arrows.

that have been accepted so far. The algorithm is based on an oracle that is input r_i and the edge weights, and outputs a lightest path p_i in $P(r_i)$. If the weight of p_i is large, then request r_i is rejected; otherwise, request r_i is routed along p_i . We refer to the online algorithm for online integral path packing by IPP. The competitive ratio of the IPP algorithm is summarized in the following theorem.

Theorem 3 ([13], following [6, 9]). *Consider an online path packing problem on an infinite graph with edge capacities such that $\inf_e c(e) \geq 1$. Assume that, for every request r_i , the length of every legal path in $P(r_i)$ is bounded by p_{\max} . Then algorithm IPP is $(2, \log(1 + 3 \cdot p_{\max}))$ -competitive online integral path packing algorithm. Moreover, the throughput of IPP for any request sequence is at least $1/2$ the throughput of any fractional packing for that sequence.*

Bounded Path Lengths. The load obtained by the IPP algorithm is logarithmic in the maximum path length p_{\max} . This suggests that p_{\max} should be polynomial in n . Lemma 4 states that limiting the number of store steps per packet by a polynomial in n decreases the fractional throughput only by a constant factor.

We use the following notation. Given a request sequence $R = \{r_i\}_i$, let $f^*(R)$ denote a maximum throughput fractional 1-packing of R , and let $f^*(R|p_{\max})$ denote a maximum throughput fractional 1-packing with respect to R under the constraint that each path is of length at most p_{\max} .

Lemma 4 (after [7]). *Let $p_{\max} \triangleq 2n \cdot (1 + \frac{B}{c})$. Then $|f^*(R|p_{\max})| > 0.31 \cdot |f^*(R)|$.*

3.4 Routing paths across 2-d Grids

Consider the following special case of routing in grids. Suppose that each path request has a specific source vertex which resides on either the south or the west side, and the destination is either the north or the east side (i.e., we can route to any vertex on the requested side). For $X \in \{S, W\}$ and $Y \in \{N, E\}$, let $\text{req}(X \rightarrow Y)$ denote the set of path requests whose source is in the X side and whose destination is the Y side (see Figure 1). The following claim establishes sufficient and necessary conditions for satisfying such path requests. We refer to the routing algorithm used in this case as *crossbar routing*.

Proposition 5 (Crossbar routing). *The path requests can be packed in the 2-dimensional $a \times b$ directed grid if and only if $|\text{req}(W \rightarrow E)| + |\text{req}(S \rightarrow E)| \leq a$ and $|\text{req}(W \rightarrow N)| + |\text{req}(S \rightarrow N)| \leq b$.*

Proof. The “only if” part is obvious. We now present a distributed algorithm that proves the “if” part. Without loss of generality, we may assume that $\text{req}(S \rightarrow N)$ and $\text{req}(W \rightarrow E)$ are empty. This assumption is satisfied by routing such requests along straight paths and giving them precedence over other requests. Thus

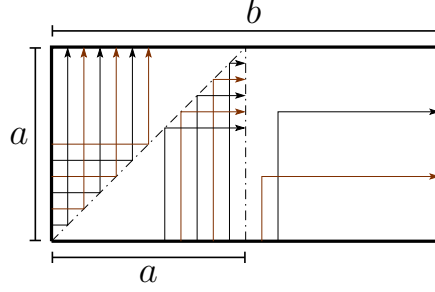


Figure 2: Satisfying the path requests in the $a \times b$ two dimensional directed grid, where $a \leq b$.

we may ignore these lines henceforth, and we are left with the task of routing $\text{req}(S \rightarrow E)$ and $\text{req}(W \rightarrow N)$ under the assumption that $|\text{req}(S \rightarrow E)| \leq a$ and $|\text{req}(W \rightarrow N)| \leq b$.

These requests are served as follows. Order the rows from bottom to top and the columns from left to right. Assume, w.l.o.g., that $a \leq b$ (the case that $a > b$ is solved analogously).

Requests whose source vertex is in the first a rows or columns turn in the vertex along the diagonal emanating from the SW corner. For example, a request in $\text{req}(W \rightarrow N)$ whose source is in row i is routed eastward for i hops, and then north for $a - i$ hops (i.e., to the north side of the grid). See Figure 2.

The requests whose source vertex is in the last $b - a$ columns are routed northward until they reach a vertex that does not receive an east-bound path from its west neighbor. Once such a vertex is found, the path turns east and continues straight until it reaches the east side of the grid. Indeed, such a right turn is always possible because $a \geq |\text{req}(S \rightarrow E)|$ and hence a “vacant row” is always found. \square

Remark 6. *Proposition 5 extends to the case of capacitated edges assuming all horizontal edges have the same capacity and all vertical edges have the same capacity. In this case, the requests can be routed iff the number of requests for each destination side is bounded by total capacity of edges crossing that side.*

4 The Packet Routing Algorithm

We now present the routing algorithm. Pseudo-code is provided in Algorithm 1. The algorithm works as follows. First, in lines 1-3, an initial filtering of the requests removes requests if too many requests originate in the same space-time vertex (see Definition 9). Then each remaining new request is processed. In lines 4-6, it is classified as either *Near* or *Far*, based on its source-destination distance (see paragraph on packet classification). *Near* requests are routed by the ROUTE-NEAR algorithm, described in Section 4.4. Each *Far* request is associated with the tiling T_j in which its source vertex belongs to a south-west quadrant of a tile. Each tiling is processed separately by three procedures: (i) The IPP algorithm, which performs online path packing over the sketch graph S_j (line 7). The outcome sketch_i is either “REJECT” or a path in a sketch graph S_j , i.e., a sequence of tiles from the initial tile to the destination tile. (ii) The INITIAL-ROUTE procedure looks for a routing within the SW-quadrant of the first tile of r_i ; its outcome is either such a path denoted init_i or “REJECT”. Only IPP and INITIAL-ROUTE may reject a far request. If both procedures are successful, then DETAILED-ROUTE is called (line 8). Detailed routing computes a path in the space-time graph, i.e., a complete schedule for each packet. In our algorithm, the sketch path for each accepted request is computed once and it is fixed, but the future part of a detailed route of a request may change due to the insertion of new packets. Therefore, the procedure DETAILED-ROUTE not only computes a path for r_i in

Algorithm 1 Top-level algorithm for packet routing in the 1-dimensional grid. Code for step t .

```

1: Let  $R_t$  be a list of new requests, sorted by source-destination distance.
2: For each vertex  $v$ , let  $R'_t(v)$  the first  $B' + c'$  requests in  $R_t$  whose source is  $v$ . ▷ filter requests
3: for each request  $r_i \in \bigcup_v R'_t(v)$  do
4:   if  $r_i \in \text{Near}$  then ROUTE-NEAR( $r_i$ )
5:   else
6:     Let  $j \in \{1, \dots, 4\}$  be s.t.  $r_i \in \text{SW}_j$  ▷ classify  $r_i$ 
7:      $\text{sketch}_i \leftarrow \text{IPP}(S_j, \text{accepted}_j, r_i)$  ▷ path lengths bounded by  $p_{\max}$ 
8:      $\text{init}_i \leftarrow \text{INITIAL-ROUTE}(\text{accepted}_j, r_i)$ 
9:     if  $\text{sketch}_i \neq \text{REJECT}$  and  $\text{init}_i \neq \text{REJECT}$  then
10:      add  $r_i$  to  $\text{accepted}_j$ 
11:      DETAILED-ROUTE( $r_i, \text{init}_i, \text{sketch}_i$ ) ▷ update routes
12:     else Reject  $r_i$ 
13:   end if
14: end if
15: end for

```

G^{st} , but may also alter the detailed routes of other requests (without changing the high-level sketch-graph routes).

An important property of IPP and INITIAL-ROUTE is that their state is determined by the requests that are actually in the system, i.e., accepted by both. (Rejected requests by either do not affect the state of the system.) This property enables us to employ the combination technique of [15]. The listing emphasizes this property by explicitly managing the sets of accepted requests for each class (denoted by accepted_j). These sets are arguments of IPP and INITIAL-ROUTE and determine their states. We now proceed to explain the algorithm in detail.

Packet classification. A request $r_i = (a_i, b_i, t_i)$ is called *near* if $b_i - a_i \leq \ell_v$, and *far* otherwise. We denote the sets of near and far requests by Near and Far , respectively. The far requests are further classified into four classes denoted by Far_j , where $\text{Far}_j \triangleq \text{Far} \cap \text{SW}_j$. Namely, Far_j is the set of far requests whose source node is in the SW-quadrant of a tile s in the tiling T_j .

Link Multiplexing. After classification, there are five classes of requests: one for Near and one for each tiling. Routes for each class are computed independently and the final result is the union of routes over all classes. This is made possible by partitioning the capacity of each edge in the space-time grid into virtual *tracks*, one track per class. The capacity of each track is $\frac{1}{5}$ of the capacity of the edge, rounded down, i.e., track capacities are $B' \triangleq \lfloor B/5 \rfloor$ and $c' \triangleq \lfloor c/5 \rfloor$. (This explains why we require that $B, c \geq 5$.)

Tiling Parameters. Tile side lengths are set so that the trivial greedy routing algorithm is $O(\log n)$ -competitive for requests that can be satisfied within a tile. Each tile has length ℓ_h and height ℓ_v , defined as follows. Recall that the maximum path length $p_{\max} = 2n \cdot (1 + \frac{B}{c})$ (cf. Lemma 4).

Definition 7. We use the following parameters.

- $k \triangleq \log(1 + 3 \cdot p_{\max})$
- $\ell_h = \lceil \frac{6k}{5c'} \rceil$ and $\ell_v = \lceil \frac{6k}{5B'} \rceil$

We summarize with the following claim.

Proposition 8. *If B/c is bounded by a polynomial in n , then the tiling parameters satisfy the following properties.*

1. $\ell_h + \ell_v = O(\log n)$.
2. *The sum of the edge capacities along each tile side is $\Theta(k)$.*
3. *For each track, the sum of the track capacities along a tile side is at least $6k$.*

Proof. Clearly $\ell_h + \ell_v = O(k)$. If B/c is polynomial in n , then $k = O(\log n)$. The sum of the edge capacities along a vertical side is $\ell_v \cdot B = \Theta(k)$. The sum of the track capacities crossing a vertical side of a tile is at least $\ell_v \cdot 5B' \geq 6k$. The capacities along a horizontal edge is bounded similarly. \square

Filtering superfluous simultaneous requests with identical sources. Since we do not impose any restriction on the requests, it could well be that many requests arrive at the same source vertex in a single time step. To deal with that, we use that fact that for each node v and step t , no more than $c + B$ requests can leave (v, t) in *any* routing. The partition of link capacities for tracks imposes a stricter limitation in the sense that within each class, no more than $c' + B'$ paths can have the same source vertex.

Definition 9. *Given a sequence R of requests, let R' denote the subsequence of R defined as follows. For each source vertex (a_i, t_i) , choose $c' + B'$ packets whose destination is closest to the source node. (If at most $B' + c'$ requests originate at the same node, then all of them are kept in R' .)*

Proposition 14 shows that rejecting the requests in $R \setminus R'$ reduces the fractional optimal throughput only by a constant factor.

4.1 Routing Rules

The routing at the high level (sketch path) is determined by the IPP algorithm. We now explain the ideas behind refining these rough paths (in the sketch graph) into actual paths (in the space-time grid). Throughout this section we consider, w.l.o.g., a single tiling T_j .

Fix a tile s IN t_j . We distinguish between the following three types of requests in Far_j (we deal with Near requests in Section 4.4).

- *Initial requests:* requests whose source vertex is in the south-west quadrant of the tile s .
- *Traversing requests:* these are requests that enter s from a specific vertex on one side (either west or south) and must leave through any vertex of another side (either east or north). The entry vertex is determined by a previously-invoked detailed routing, and the exit side is determined by the sketch path.
- *Final requests:* these are requests whose sketch path ends in tile s . The destination of a final request is the north side of s .¹

Each tile is partitioned into 4 quadrants, denoted NE, SE, SW and NW. We constrain the way requests are routed within a tile using the following rules (see Figure 3; no request may cross a thick line).

1. Initial requests always start in the SW-quadrant and are routed to the north or east side of the SW-quadrant along a straight path. The SW-quadrant of each tile is reserved for routing of initial requests.
2. Traversing requests whose source and destination sides are opposite (e.g., from the south to the north side) are routed along a straight path.

¹The detailed path for r_i should end in a copy of the destination vertex b_i in s . As any path that reaches the north side of s also reaches a copy of b_i , we use the pessimistic assumption that $row(b_i)$ is the northmost row of tile s .

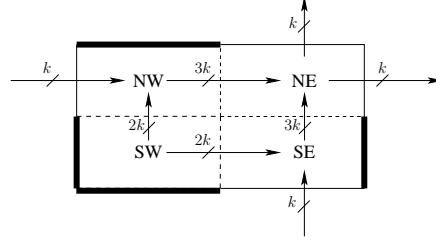


Figure 3: The quadrants of a tile. Traversing requests may not cross the thick border lines. Upper bounds on the number of paths in each class that traverse the quadrant sides follow from the path packing algorithm IPP.

3. Paths enter the tile either in the east half of the south side or the north half of the west side.
4. Paths exit the tile through either the east half of the north side or the north half of the east side.

4.2 Procedure INITIAL-ROUTE

Initial routing takes place in the SW-quadrant of the first tile of a *Far* request. The goal of INITIAL-ROUTE is modest: route the request to the boundary of the SW-quadrant. This is done greedily along straight paths if possible: it may be the case that some of the edges are already reserved for another request (also routed by INITIAL-ROUTE of the same tile). If no straight path is available, INITIAL-ROUTE returns REJECT. This means that incoming traffic (of earlier requests) continues uninterrupted along a straight path. Only remaining capacity along edges that emanate from a vertex (i, j) , if any, is used for routing the requests that originate in (i, j) .

4.3 Procedure DETAILED-ROUTE

The goal in detailed routing is to compute a detailed path p_i in the space-time graph G^{st} given a sketch path $sketch_i$ in the sketch graph S_j and the initial part of the route $init_i$. The sketch path specifies the sequence of tiles to be traversed by the detailed path. In addition, the sketch path specifies the tile sides through which the detailed path should enter and exit each tile. Requests that have been assigned a sketch path and an initial route must be successfully routed by detailed routing.

Detailed routing is computed by applying crossbar routing (cf. [Proposition 5](#)) to the NW, SE and NE quadrants. This routing is computed based on the present requests. As new requests arrive, the future portions of the detailed routes may change dynamically so that all requests which are “in progress” will reach their destination. Below we argue that crossbar routing indeed succeeds.

Claim 10. *Detailed routing successfully completes the route of each accepted far request.*

Proof. By [Proposition 5](#), to ensure successful routing it is sufficient to bound the number of paths that need to traverse a quadrant by the capacity of the quadrant side. By [Proposition 8](#), the track capacity of each quadrant side is at least $3k$. We now prove upper bounds on the number of paths that traverse each quadrant side (see [Figure 3](#)). The IPP path packing algorithm is a k -packing over the sketch graph (whose edges have unit capacity). It follows that at most k paths traverse each side of the tile. As every request that originates in the SW-quadrant of a tile must exit the tile, there are at most $2k$ paths that traverse each side of the SW-quadrant (although their sum is also bounded by $2k$). Hence the upper bounds depicted in [Figure 3](#) follow. We need to elaborate more on the NE-quadrant because it is also used for routing final requests (i.e., requests

that do not exit the tile, but do want to reach its top row). Consider the north side of the NE-quadrant. There are at most k traversing requests that wish to exit the tile. In addition, there are at most $2k$ final requests that wish to reach to top row (as each final far request must have entered the tile). Thus, in total there are at most $3k$ paths that wish to reach the top side of the NE-quadrant. To summarize, the number of paths that wish to reach any quadrant side is bounded by the side's capacity, and hence by [Proposition 5](#), detailed routing succeeds. \square

Finally, we note that in order for DETAILED-ROUTE to be well defined, we compute it in tiles in column-major order, i.e., we start with the bottom tile of the leftmost row and go up, then the bottom tile of the second-from left column and go up etc. This ensures that when we reach a tile, all input vertices are fixed. We remark that detailed routing can be executed in a local distributed manner; in each time step, each vertex needs only to know the initial paths the sketch paths of the incoming packets.

4.4 Procedure ROUTE-NEAR

Finally, we describe the algorithm for the near requests. The ROUTE-NEAR Algorithm is extremely simple: it never stores a packet (i.e., it uses only vertical edges in G^{st} , and gives precedence to older requests). In more detail, upon arrival of a request $r_i \in \text{Near}$, the algorithm checks the number of requests already routed along the outgoing vertical edge (from (a_i, t) to $(a_i + 1, t + 1)$). If this number is less than c' , then the algorithm routes r_i along the vertical path in G^{st} from (a_i, t) to $(b_i, t + (b_i - a_i))$. Note that these edges occur in the future, and hence cannot have been saturated by ROUTE-NEAR if the edge outgoing from (a_i, t) is not saturated. If there is no free capacity in the outgoing vertical edge, r_i is rejected. Note that if r_j is accepted, then it guaranteed to reach its destination.

5 Analysis of Competitive Ratio of the Routing Algorithm

Our goal is to prove the following theorem for a directed line G of n vertices with buffer sizes B and link capacities c , where $B, c \in [5, \log n]$.

Theorem 11. *Algorithm 1 is $O(\log n)$ -competitive with respect to the throughput of a maximum fractional routing.*

We translate the problem to a path packing problem over the space-time graph G^{st} . Let $f_{G^{st}}^*(R)$ denote a maximum throughput fractional routing, and let $|f_{G^{st}}^*(R)|$ denote its throughput. Let $|\text{ALG}(R)|$ denote the throughput of the online packet algorithm. [Theorem 11](#) follows directly from the following lemma.

Lemma 12. *For every sequence of requests R , $|f_{G^{st}}^*(R)| \leq O(\log n) \cdot |\text{ALG}(R)|$.*

We outline the proof of [Lemma 12](#). We scale the capacities down by a factor of $\Theta(k) = \Theta(\log n)$ in the sketch graph. By linearity, this reduces the optimal fractional throughput by the same factor (see [Proposition 13](#)). We show that the filtering stage in [Line 2](#) incurs only a constant factor reduction to the optimal fractional throughput (see [Proposition 14](#)). The filtered requests R' are partitioned into near requests and far requests (which are further partitioned into 4 classes, one per tiling). The far and near requests are analyzed separately. The analysis of the throughput for far requests builds on the competitive ratio of the IPP algorithm and the INITIAL-ROUTE algorithm (see [Claim 15](#) and [Claim 16](#)). By applying the combining analysis of Kleinberg and Tardos [[15](#)], we show that the competitive ratio for the combined algorithm is the sum of the two algorithms (see [Claim 17](#)). In [Theorem 18](#), we show that the ROUTE-NEAR algorithm succeeds in routing a logarithmic fraction of the filtered near requests. In [Section 5.4](#), the parts of the proof are combined together to prove [Lemma 12](#).

5.1 Scaling and Filtering

One advantage of working with fractional routings is that, by linearity, the throughput scales exactly with the capacities. Let $f_{S_j}^*(R)$ denote a maximum throughput fractional routing in the sketch graph S_j . Recall that the sketch graph has unit capacities. Coalescing of vertices of G^{st} in each tile results with edge capacities that are $\Theta(k) = \Theta(\log n)$. Hence, we obtain the following proposition.

Proposition 13. $|f_{S_j}^*(R)| = \frac{1}{\Theta(k)} \cdot |f_{G^{st}}^*(R)|$.

Recall that by [Definition 9](#), in the input sequence R' , at most $B' + c'$ requests originate in each space-time vertex.

Proposition 14. $|f_{G^{st}}^*(R)| \leq 9 \cdot |f_{G^{st}}^*(R')|$.

Proof. Fix a space-time vertex $v = (a_i, t_i)$. Let R_v (reps. R'_v) denote requests in R (reps. R') that originate in v . Let $f = f_{G^{st}}^*(R)$. Consider the flow $f/9$. For every vertex v , the amount of flow that originates in v is bounded by $(B + c)/9 \leq B' + c'$. Divert flow from in $f/9$ from $R_v \setminus R'_v$ to R'_v along shorter paths, to obtain a flow g with respect to R' such that $|g| = |f|/9$. Since $|f_{G^{st}}^*(R')| \geq |g|$, the proposition follows. \square

5.2 Far Requests

Two algorithms determine whether a far request is rejected: (i) the IPP path packing algorithm over the sketch graph, and (ii) the INITIAL-ROUTE algorithm that deals with routing in the initial SW-quadrant of the source tile. We begin by showing that, if invoked separately, each of these algorithms accepts at least a constant fraction of the maximum fractional throughput over the sketch graph.

Let R'_j denote the subsequence of requests in R' that are in the class Far_j . Suppose we invoke the IPP algorithm in isolation over the sketch graph S_j with the input sequence R'_j . By isolation we mean that the accepted requests are determined solely by IPP. Let $|IPP_{S_j}(R'_j)|$ denote the number of requests that are accepted by this invocation.

Claim 15. $|f_{S_j}^*(R'_j)| \leq \frac{2}{0.31} \cdot |IPP_{S_j}(R'_j)|$.

Proof. By [Lemma 4](#), the restriction of the path lengths by p_{\max} only reduces the fractional throughput by a factor less than 0.31. By [Theorem 3](#), the IPP algorithm is $(2, k)$ -competitive, and hence its throughput is half the optimal fractional throughput with bounded path lengths. \square

Let $|INITIAL-ROUTE(R'_j)|$ denote the number of requests that are accepted by INITIAL-ROUTE if invoked in isolation with the input sequence R'_j .

Claim 16. $|f_{S_j}^*(R'_j)| \leq 2 \cdot |INITIAL-ROUTE(R'_j)|$.

Proof. A far request must exit the tile in which it begins. The edge capacities in the sketch graph are unit. Hence, the amount of flow in $f_{S_j}^*(R'_j)$ that originates in each tile is at most 2. On the other hand, if a positive amount of flow originates in a tile s , then at least one request starts in the SW-quadrant of s . Hence INITIAL-ROUTE(R'_j) accepts at least one request that begins in s . \square

A naïve analysis of the requests accepted by the conjunction of the IPP and INITIAL-ROUTE algorithms implies that the accepted requests are in the intersection, which might be empty. However, in our algorithm the subsequence of accepted requests is determined by both algorithms, and this set of accepted requests determines the state of both algorithms. Hence, by applying the combining analysis of Kleinberg and Tardos [\[15\]](#), the combined competitive ratio is shown to be the sum of the isolated competitive ratios.

Claim 17. $f_{S_j}^*(R'_j) \leq (2 + \frac{2}{0.31}) \cdot |\text{ALG}(R'_j)|$.

Proof. To simplify notation, let A_1 denote the IPP algorithm in isolation, A_2 denote the INITIAL-ROUTE algorithm in isolation, and ALG denote the combined algorithm. Let I denote the input sequence R'_j .

Consider the execution of ALG with the input sequence I . Let X denote the subsequence of requests accepted by the combined algorithm A on input I (i.e., $X = A(I)$). Let X_j denote the subsequence of requests accepted by algorithm A_j during this same execution. Note that $X = X_1 \cap X_2$. Let us rewrite $f_{S_j}^*(I)$ as a function $f^* : I \rightarrow [0, 1]$ in which $f^*(i)$ equals the amount of flow assigned to request $i \in I$ by $f_{S_j}^*(I)$. We abuse notation and view X also as its characteristic function (i.e., $X(i) = 1$ if and only if $i \in X$).

Consider the packet routing version in which requests have demands in $[0, 1]$. A demand of 1 corresponds to the situation till now. A demand of zero means that the request does not appear (the input skips over this request). A fractional demand means that (i) at most this fraction can be routed by the fractional routing, and (ii) the request only occupies this fraction of the capacity of an edge. Fractional demand functions can be added and multiplied. If both functions are integral, then addition corresponds to the union, and multiplication corresponds to the intersection.

Consider the demand function $Z_\ell \triangleq X + f^* \cdot (1 - A_\ell)$, for $\ell \in \{1, 2\}$. First, note that it attains values in $[0, 1]$ because (i) an accepted request $i \in X$ satisfies $A_\ell(i) = 1$, and (ii) a rejected request $i \notin X$ satisfies $X(i) = 0$. Note also that $f^* \cdot (X + 1 - A_\ell)$ is a feasible fractional flow with respect to the demand Z_ℓ , hence

$$|f_{S_j}^*(Z_\ell)| \geq \|f^* \cdot (X + 1 - A_\ell)\|_1. \quad (1)$$

We claim that the isolated algorithm A_ℓ on input Z_ℓ accepts exactly X . To prove this consider the sequence of states of $A_\ell(Z_\ell)$ and $A(X)$. Let S_ℓ^i denote the prefix of accepted requests by $A_\ell(Z_\ell)$ till but not including the arrival of request i . Similarly, let X^i denote the prefix of accepted requests by $A(X)$ till but not including the arrival of request i . If $S_\ell^i = X^i$, for every i , then $A_\ell(Z_\ell) = X$. Now we prove that $S_\ell^i = X^i$ by induction on i . Before the first request, both sets are empty. The induction step for $i \in X$ is easy because A_ℓ accepts i when the state is X^i . On the other hand, if $i \notin X$ and $Z_\ell(i) > 0$, then $A_\ell(i) = 0$.

Let CR_ℓ denote the competitive ratio of A_ℓ . Then

$$|f_{S_j}^*(Z_\ell)| \leq CR_\ell \cdot A_\ell(Z_\ell) = CR_\ell \cdot |X| \quad (2)$$

By Equations 1 and 2

$$\|f^* \cdot (X + 1 - A_\ell)\|_1 \leq CR_\ell \cdot |X|. \quad (3)$$

Observe that $1 \leq (X + 1 - A_1) + (X + 1 - A_2)$. Indeed, if $i \in X$, then $A_1(i) = A_2(i) = 1$, so both sides equal 1. If $i \notin X$, then $A_1(i)$ or $A_2(i)$ equals zero (perhaps both), and hence the right hand side is at least 1. Hence,

$$\begin{aligned} \|f^*\|_1 &\leq \|f^* \cdot (X + 1 - A_1)\|_1 + \|f^* \cdot (X + 1 - A_2)\|_1 \\ &\leq CR_1 \cdot |X| + CR_2 \cdot |X|, \end{aligned}$$

and the claim follows. \square

5.3 Near Requests

In this section we analyze the competitive ratio of the ROUTE-NEAR algorithm with respect to near requests. Recall that: (1) A request is a near request if the distance from the source to the destination is at most ℓ_v . Note that $\ell_v = \Theta(\frac{\log n}{B})$ and $B < \log(1 + 3p_{\max}) = O(\log n)$. (2) The incoming requests are filtered so that at most $B' + c'$ requests originate in every space-time vertex.

The following theorem states that ROUTE-NEAR succeeds in routing at least a logarithmic fraction of the filtered near requests. This theorem implies that the throughput is at least a logarithmic fraction of the optimal fractional routing of the filtered near requests.

Theorem 18. $|\text{ALG}(\text{Near})| \geq \Omega(\frac{1}{\log n}) \cdot |\text{Near}|$.

Proof. It suffices to prove that $|\text{ALG}(\text{Near})| \geq \Omega(\frac{1}{\log n}) \cdot |\text{Near} \setminus \text{ALG}(\text{Near})|$. Consider the following bipartite conflict graph. Nodes on side L are the requests of $\text{ALG}(\text{Near})$, and nodes on side R are the requests of $\text{Near} \setminus \text{ALG}(\text{Near})$. There is an edge $(r_i, r_j) \in L \times R$ if r_j is rejected by the ROUTE-NEAR Algorithm and the vertical route of r_i traverses the source vertex (a_j, t_j) of r_j . A request $r_i \in L$ conflicts with at most $B' + c'$ requests in each vertex. Hence, the degree of r_i in the conflict graph is at most $(B' + c') \cdot \ell_v$. On the other hand, the degree of $r_j \in R$ equals c' (where c' is the capacity of the track reserved for the near requests). By counting edges on each side we conclude that

$$\sum_{r_i \in L} \deg(r_i) = \sum_{r_j \in R} \deg(r_j).$$

Hence,

$$(B' + c') \cdot \ell_v \cdot |L| \geq c' \cdot |R|.$$

We conclude that

$$|L| \geq \frac{c'}{(B' + c')\ell_v} \cdot |R|.$$

As $\frac{(B' + c')\ell_v}{c'} = \Theta(\frac{\log n}{c'} + \frac{\log n}{B'}) = O(\log n)$, and the theorem follows. \square

5.4 Putting Things Together

In this section we prove [Lemma 12](#). We partition the input sequence R' into Near and R'_j , for $j \in \{1, 2, 3, 4\}$ (recall that $R'_j = R' \cap \text{Far}_j$). By subadditivity,

$$|f_{G^{st}}^*(R')| \leq |f_{G^{st}}^*(\text{Near})| + \sum_{j=1}^4 |f_{G^{st}}^*(R'_j)|, \quad (4)$$

$$|\text{ALG}(R')| = |\text{ALG}(\text{Near})| + \sum_{j=1}^4 |\text{ALG}(R'_j)|. \quad (5)$$

In order to bound the ratio $|f_{G^{st}}^*(R')|/|\text{ALG}(R')|$, it suffices to separately bound the ratios of the terms. Indeed, by [Theorem 18](#)

$$|f_{G^{st}}^*(\text{Near})| \leq O(\log n) \cdot |\text{ALG}(\text{Near})|.$$

By [Proposition 13](#) and [Claim 17](#),

$$|f_{G^{st}}^*(R'_j)| \leq O(\log n) \cdot |\text{ALG}(R'_j)|.$$

Finally, by [Proposition 14](#), $|f_{G^{st}}^*(R)| \leq 9 \cdot |f_{G^{st}}^*(R')|$. Since $\text{ALG}(R) = \text{ALG}(R')$, the lemma follows.

6 Extension to d -Dimensional Grids

The following theorem is proved by extending [Algorithm 1](#) for a line network to d -dimensional grid.

Theorem 19. *For $B, c \in [2^{d+1}+1, \log n]$, there is a deterministic $2^{O(d)} \cdot \log^d n$ -competitive online algorithm for the throughput maximization problem.*

(*sketch*). As in the one-dimensional case, perform a space-time transformation on the d -dimensional n -node grid G to obtain the $(d+1)$ -dimensional space-time grid G^{st} . Partition G^{st} to $\ell_1 \times \dots \times \ell_{d+1}$ subgrids (or subcubes). The side length of a subgrid equals ℓ_v for directions that correspond to forward steps and ℓ_h in the direction that corresponds to store steps. There are two offsets per dimension, resulting with 2^{d+1} tilings. The number of tracks equals the number of offsets plus one (the extra one is for the near requests), hence we require that $B, c \geq 2^{d+1} + 1$. Similarly to the 1-dimensional case, a request is classified as a near request if the distance from the source to the destination is at most $d \cdot \ell_v$. Detailed routing within a tile is successful by the following observation. Every time a packet cannot turn to the direction that is dictated by its sketch path, there is a packet that did turn to its desired direction. Since the number of path emanating from each tile is bounded by the quadrant-side capacity, we conclude that every packet will eventually turn, if needed, within its quadrant, thus respecting its sketch path.

Since the link capacity to track capacity ratio is $O(2^d)$, this scaling of capacities incurs an $O(2^d)$ factor to the competitive ratio. The sketch graph is obtained in the same way, with the exception that edge capacities are set to $\frac{1}{d+1}$ (so that that the number of paths that IPP routes out of a $(d+1)$ -dimensional tile is at most $O(\log n)$). The ratio of edge capacities in G^{st} between adjacent faces of tiles and the capacity of the edge in the sketch graph is $O(\log n)^d$. This incurs an additional factor of $O(\log n)^d$ for routing far requests due to capacity scaling. The routing of near requests succeeds in routing at least a fraction of $d \cdot \log n$ of the near requests. We conclude that the competitive ratio is determined by the far requests, and hence the theorem follows. \square

7 Conclusion

In this paper we presented an online deterministic packet routing algorithm. For the one dimensional grid (with constant-size buffers and constant-capacity links), this algorithm closes the gap with the best throughput achieved by a randomized algorithm. This closes a problem which was open for more than a decade, but still leaves open quite a few problems. The most urgent one is to reduce the gap between the upper and lower bounds on the competitive ratio. Currently the best upper bound is $O(\log n)$ for the line, and we are not aware of any non-trivial lower bound. We note that reducing the upper bound to $o(\log n)$ seems to require new techniques, as the reduction to online path packing introduces a logarithmic factor in the competitive ratio.

Another important question is to come up with reasonable distributed algorithms. Even though, as mentioned above, the SDN model shifts many network operation tasks to the centralized setting, it is very interesting to find out what can be done without a central coordinator.

References

- [1] M. Adler, S. Khanna, R. Rajaraman, and A. Rosén. Time-constrained scheduling of weighted packets on trees and meshes. *Algorithmica*, 36(2):123–152, 2003.
- [2] M. Adler, A. L. Rosenberg, R. K. Sitaraman, and W. Unger. Scheduling time-constrained communication in linear networks. *Theory Comput. Syst.*, 35(6):599–623, 2002.
- [3] W. Aiello, E. Kushilevitz, R. Ostrovsky, and A. Rosén. Dynamic routing on networks with fixed-size buffers. In *SODA*, pages 771–780, 2003.
- [4] S. Angelov, S. Khanna, and K. Kunal. The network as a storage device: Dynamic routing with bounded buffers. *Algorithmica*, 55(1):71–94, 2009. (Appeared in APPROX-05).
- [5] B. Awerbuch, Y. Azar, and A. Fiat. Packet routing via min-cost circuit routing. In *ISTCS*, pages 37–42, 1996.
- [6] B. Awerbuch, Y. Azar, and S. Plotkin. Throughput-competitive on-line routing. In *FOCS '93: Proceedings of the 1993 IEEE 34th Annual Foundations of Computer Science*, pages 32–40, Washington, DC, USA, 1993. IEEE Computer Society.
- [7] Y. Azar and R. Zachut. Packet routing and information gathering in lines, rings and trees. In *ESA*, pages 484–495, 2005. (See also manuscript in <http://www.cs.tau.ac.il/~azar/>).
- [8] Y. Bartal and S. Leonardi. On-line routing in all-optical networks. In *ICALP*, pages 516–526, 1997.
- [9] N. Buchbinder and J. S. Naor. Improved bounds for online routing and packing via a primal-dual approach. *Foundations of Computer Science, Annual IEEE Symposium on*, 0:293–304, 2006.
- [10] N. Buchbinder and J. S. Naor. The design of competitive online algorithms via a primal-dual approach. *Foundations and Trends in Theoretical Computer Science*, 3(2-3):99–263, 2009.
- [11] G. Even and M. Medina. An $O(\log n)$ -Competitive Online Centralized Randomized Packet-Routing Algorithm for Lines. In *ICALP (2)*, pages 139–150, 2010.
- [12] G. Even and M. Medina. Online packet-routing in grids with bounded buffers. In *Proc. 23rd Ann. ACM Symp. on Parallelism in Algorithms and Architectures (SPAA)*, pages 215–224, 2011.
- [13] G. Even and M. Medina. Online packet-routing in grids with bounded buffers. *CoRR*, abs/1407.4498, 2014.
- [14] N. Gude, T. Koponen, J. Pettit, B. Pfaff, M. Casado, N. McKeown, and S. Shenker. Nox: Towards an operating system for networks. *SIGCOMM Comput. Commun. Rev.*, 38(3):105–110, July 2008.
- [15] J. M. Kleinberg and É. Tardos. Disjoint paths in densely embedded graphs. In *FOCS*, pages 52–61, 1995. (See also manuscript in <http://www.cs.cornell.edu/home/kleinber/>).
- [16] H. Räcke and A. Rosén. Approximation algorithms for time-constrained scheduling on line networks. In *SPAA*, pages 337–346, 2009.
- [17] J. S. Turner. Strong performance guarantees for asynchronous buffered crossbar scheduler. *IEEE/ACM Trans. Netw.*, 17(4):1017–1028, 2009.

A $\sqrt{\log n}$ -competitiveness of initial routing

Lemma 20. *Fix a tile s and let q denote its SW quadrant. Suppose that the sources of m path requests are in q . Then $\Omega(\sqrt{m})$ path requests are served by the initial routing in q .*

Proof. We restrict attention to rows of q which contain at least B' sources and columns of q which contain at least c' sources of requests which cannot be routed horizontally. Since all other packets are trivially routed by the algorithm, we may assume w.l.o.g. that there are no other packets with sources in q .

Let y denote the number of rows that contain a source vertex of an initial request, and let x denote the number of columns that contain a source vertex of an initial request that is not routed horizontally. Clearly, $m \leq xy(B' + c')$. On the other hand, detailed routing in q serves $yB' + xc'$ requests. We now prove that $yB' + xc' = \Omega(\sqrt{xy(B' + c')})$.

Without loss of generality, assume that $c' \geq B'$. Thus it suffices to prove that

$$\sqrt{\frac{y}{x}} \cdot B' + \sqrt{\frac{x}{y}} \cdot c' = \Omega(\sqrt{c'}).$$

We proceed with case analysis. If $y \leq x$, then $\sqrt{\frac{y}{x}} \cdot B' + \sqrt{\frac{x}{y}} \cdot c' \geq c'$, as required. Otherwise $y > x$. We further distinguish between two cases:

1. If $y/x \geq c'/B'$, then $\sqrt{\frac{y}{x}} \cdot B' + \sqrt{\frac{x}{y}} \cdot c' \geq \sqrt{\frac{c'}{B'}} \cdot B' \geq \sqrt{c'}$, as required.
2. If $x/y > B'/c'$, then $\sqrt{\frac{y}{x}} \cdot B' + \sqrt{\frac{x}{y}} \cdot c' \geq \sqrt{\frac{B'}{c'}} \cdot c' \geq \sqrt{c'}$, as required.

□

Note that, if a single requested is input to a SW-quadrant, then initial routing accepts it.